Optimal Designs for 2^k Factorial Experiments with Binary Response

Jie Yang[1], Abhyuday Mandal[2] and Dibyen Majumdar[1]
[1]University of Illinois at Chicago and [2]University of Georgia

April 9, 2012

Abstract:

We consider the problem of obtaining locally D-optimal designs for factorial experiments with binary response and k qualitative factors at two levels each. Yang, Mandal and Majumdar (2011) considered this problem for 2^2 factorial experiments. In this paper, we generalize the results for 2^k designs and explore in new directions. We obtain a characterization for a design to be locally D-optimal. Based on this characterization, we develop efficient numerical techniques to search for locally D-optimal designs. We also investigate the properties of fractional factorial designs and study the robustness, with respect to the initial parameter values of locally D-optimal designs. Using prior distribution on the parameters, we investigate EW D-optimal designs, that are designs which maximize the determinant of the expected information matrix. It turns out that these designs are much easier to find and still highly efficient compared to Bayesian D-optimal designs, as well as quite robust.

Key words and phrases: Generalized linear model, full factorial design, fractional factorial design, D-optimality, uniform design, EW D-optimal design

1 Introduction

The goal of many scientific and industrial experiments is to investigate the effect of several factors on a response that is binary. Our objective is to explore optimal and efficient designs for these experiments. In Yang, Mandal and Majumdar (2011) we considered this problem for the 2^2 factorial experiment. In this paper, we consider the general 2^k case, that is, experiments with k factors at two levels each. The factors are either qualitative, or quantitative with only two levels possible. In other words,

the experimenter lacks the option of adjusting the levels of each factor beyond the two fixed ones.

Instances of experiments of the sort we are interested in abound in different areas of application. Miller and Sitter (2001) described an experiment with 12 runs that investigated ways of reducing the level of toxic contaminant (stripper) from the waste stream of a chemical operation. The experiment considered nine factors at two levels each. Grimshaw et al. (2001) described an experiment on the vision-based targeting component of an automatic car-refueling system that was performed at Brigham Young University in 1998. The objective of the experiment was to find out if the robotic system could consistently insert the nozzle at the end of a telescoping arm into a gas inlet pipe or not. This experiment had ten factors at two levels each. Seo, Goldschmidt-Clermont and West (2007) discussed an experiment where the response could be whether or not a particular symptom is developed in mice. This study investigated three key environmental factors, age, gender and dietary fat intake along with a key genetic factor related to the ApoE (Apolipoprotein E) gene pathway. Hamada and Nelder (1997) discussed a 2^{4-1} fractional factorial experiment performed at IIT Thompson laboratory that was originally reported by Martin, Parker and Zenick (1987). This was a windshield molding slugging experiment where the outcome was whether the molding was good or not. There were four factors each at two levels: (A) poly-film thickness (0.0025, 0.00175), (B) oil mixture ratio (1:20, 1:10), (C) material of gloves (cotton, nylon), and (D) the condition of metal blanks (dry underside, oily underside). Vasandani and Govindaraj (1995) considered knowledge organization in intelligent tutoring systems where the objective was to determine whether a system fails or not. Street and Burgess (2007) reported the study conducted by Severin (2000) where she investigated six attributes (pizza type, type of crust, ingredients, size, price and delivery time) each at two levels to examine whether they make take-out pizza outlets more attractive or not. Nair *et al.* (2008) reported a 2^{6-2} fractional factorial experiment used in the study called "Project Quit", a Center for Health Communications Research project on smoking cessation. They studied six factors each at two levels and the primary outcome measure was abstinence over a 7-day period assessed by the question, "Did you smoke a tobacco cigarette in the past 7 days?"

In many real-life applications, interest lies in accessing a situation with a binary outcome (*e.g.*, in drug discovery, whether a compound is toxic or not, or in car manufacturing, whether a sensor works or not) but the practitioners often use a surrogate continuous response in order to use the optimal designs obtained under linear models.

We assume that the process under study is adequately described by a generalized linear model (GLM). GLMs have been widely used for modeling binary response. Stufken and Yang (2012) noted that "the study of optimal designs for experiments that plan to use a GLM is however not nearly as well developed (see also Khuri,

Mukherjee, Sinha and Ghosh, 2006), and tends to be much more difficult than the corresponding and better studied problem for the special case of linear models."

The optimal designs are obtained using the D-optimality criterion that maximizes the determinant of the information matrix. In order to overcome the difficulty posed by the dependence of the design optimality criterion on the unknown parameters, we primarily use the local optimality approach of Chernoff (1953) in which the parameters are replaced by assumed values. We refer the reader to the paper by Khuri, Mukherjee, Sinha, and Ghosh (2006) for details of the theory of designs under generalized linear models.

We obtain theoretical results and develop algorithms for obtaining locally optimal designs. Experiments with factors at two levels each are often used in screening experiments, where the objective is to determine which factors among a large set are effective. Although we explore designs for "full factorials", i.e., ones in which observations are taken at every possible level combination, when the number of factors is large, full factorials are practically infeasible. Hence the study of "fractional factorial" designs occupies a substantial part of the linear-model based design literature, and we too study these designs in our generalized linear model setup.

A natural question that arises when we use local optimality is whether the optimal designs are robust. The study of robustness with respect to the initial parameter values of locally D-optimal designs, therefore, is a significant part of our focus.

For situations where reliable local values of the parameters are difficult to obtain but the experimenter may be able to specify a prior distribution, we suggest *EW optimal* designs, where the information matrix is replaced by its expectation under the prior. This is one of the suggested alternatives to formal Bayes optimality in Atkinson, Donev, and Tobias (2007). It has been used by Zayats and Steinberg (2010) for optimal designs for detection capability of networks. Effectively this reduces to a locally optimal design with local values of the "weight" parameters replaced by their expectations. It turns out that the EW optimal designs have strong efficiency and robustness characteristics as local designs, and in addition, they are very good and easy-to-compute surrogates for Bayes optimal designs.

Beyond theoretical results, the question that may be asked is whether these results give the user any advantage in real experiments. Would the experimenter gain much by using these results instead of designs that are optimal for the linear model, for instance, uniform complete designs or regular fractions. After all, on 2^2 experiments, Yang, Mandal and Majumdar (2011) concluded that the uniform design (i.e., full factorial) had many desirable properties. In the more complex and more realistic setup of 2^k experiments considered in the present work it turns out that the answer to the question posed above is a definite "yes". In most situations we gain considerably by taking advantage of the results of this paper instead of using standard linear-model results. We briefly outline two conclusions that may be derived from our work.

First, unlike the linear model case, all regular fractions are not equivalent. Indeed, if we have some knowledge of the prior parameters, we will be able to use this to identify a fractional factorial design that may not be a regular fraction. Second, even though in the absence of any knowledge of the parameters, an uniform design may be the best, even with a little prior information, we will be able to choose an EW optimal design that is more robust with respect to the initial parameter values of locally D-optimal designs, in terms of minimizing the *maximum relative loss*.

As mentioned in Yang, Mandal and Majumdar (2011), the paper by Graßhoff and Schwabe (2008) has some relevant results for the $k = 2$ factor case. For the general case, González-Dávila, Dorta-Guerra and Ginebra (2007), and Dorta-Guerra, González-Dávila and Ginebra (2008) obtained some results for 2^k experiments with binary response. They obtained an expression for the D-criterion and studied several special cases.

This paper is organized as follows. In Section 2 we describe the preliminary setup. In Section 3 we provide several results for the locally D-optimal designs including the uniqueness of the D-optimal designs, characterization for a design to be locally D-optimal, the concept of EW D-optimal designs, and algorithms for searching D-optimal designs. In Section 4 we discuss the properties of fractional factorial designs. We address the robustness of D-optimal designs in Section 5 and conclude with some examples and remarks in Section 6. Proofs and some details about the algorithms are relegated to the Appendix and the Supplementary Materials of this paper.

2 Preliminary Setup

Consider a 2^k experiment with binary response, i.e., an experiment with k explanatory variables at 2 levels each. Suppose n_i units are allocated to the ith experimental condition such that $n_i \geqslant 0$, $i = 1, \ldots, 2^k$, and $n_1 + \cdots + n_{2^k} = n$. We suppose that n is fixed and the problem is to determine the "optimal" n_i's. In fact, we write our optimality criterion in terms of the proportions:

$$p_i = n_i/n, \ i = 1, \ldots, 2^k$$

and determine the "optimal" p_i's over real values. Since n_i's are integers, an optimal design obtained in this fashion may not always be "feasible". In Section 3.3.2 we will consider the design problem over integer n_i's.

Suppose η is the linear predictor that involves main effects and interactions which are assumed to be in the model. For instance, for a 2^3 experiment with a model that includes the main effects and the two-factor interaction of factors 1 and 2, $\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2$, where each $x_i \in \{-1, 1\}$. The aim of the experiment is to obtain inferences about the parameter vector of factor effects $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_{12})'$. In the framework of generalized linear models, the expectation of the response Y,

$E(Y) = \pi$, is connected to the linear predictor η by the link function g: $\eta = g(\pi)$ (McCullagh and Nelder (1989)). For a binary response, the commonly used link functions include logit, probit, log-log, and complimentary log-log links.

The maximum likelihood estimator of $\boldsymbol{\beta}$ has an asymptotic covariance matrix (McCullagh and Nelder, 1989; Khuri, Mukherjee, Sinha and Ghosh, 2006) that is the inverse of $nX'WX$, where $W = \text{diag}\{w_1p_1, ..., w_{2^k}p_{2^k}\}$, $w_i = \left(\frac{d\pi_i}{d\eta_i}\right)^2 / (\pi_i(1-\pi_i)) \geq 0$, η_i and π_i correspond to the ith experimental condition for η and π, and X is the "design matrix". For example, for a full factorial 2^3 experiment with model $\eta = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_{12}x_1x_2$,

$$X = \begin{pmatrix} +1 & +1 & +1 & +1 & +1 \\ +1 & +1 & +1 & -1 & +1 \\ +1 & +1 & -1 & +1 & -1 \\ +1 & +1 & -1 & -1 & -1 \\ +1 & -1 & +1 & +1 & -1 \\ +1 & -1 & +1 & -1 & -1 \\ +1 & -1 & -1 & +1 & +1 \\ +1 & -1 & -1 & -1 & +1 \end{pmatrix} \tag{1}$$

Note that the matrix $X'WX$ may be viewed as the *per-observation* information matrix. A D-optimal design maximizing $|X'WX|$ depends on the w_i's, which depend on the regression parameters $\boldsymbol{\beta}$ and the link function g. In this paper, we discuss D-optimal designs in terms of w_i's so that our results can be applied to all link functions.

3 Locally D-Optimal Designs

In this section, we consider locally D-optimal designs for the full factorial 2^k experiment. The goal is to find an optimal $\mathbf{p} = (p_1, p_2, \ldots, p_{2^k})'$ which maximizes $f(\mathbf{p}) = |X'WX|$ for specified values of $w_i \geq 0, i = 1, \ldots, 2^k$. Here $p_i \geq 0$, $i = 1, \ldots, 2^k$ and $\sum_{i=1}^{2^k} p_i = 1$. It is easy to see that there always exists a D-optimal allocation $\mathbf{p}$ since the set of all feasible allocations is bounded and closed. On the other hand, the uniqueness of D-optimal designs is usually not guaranteed (see Remark 1).

3.1 Characterization of locally D-optimal designs

Suppose the parameters (main effects and interactions) are $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_d)'$, where $d \geq k$. The following lemma expresses our objective function as an order-(d+1) homogeneous polynomial of $p_1, \ldots, p_{2^k}$, which is useful in deriving optimal p_i's.

Lemma 3.1 *Let $X[i_1, i_2, \ldots, i_{d+1}]$ be the $(d+1) \times (d+1)$ sub-matrix consisting of the i_1th, i_2th, ..., i_{d+1}th rows of the design matrix X. Then*

$$f(\mathbf{p}) = |X'WX| = \sum_{1 \le i_1 < i_2 < \cdots < i_{d+1} \le 2^k} |X[i_1, i_2, \ldots, i_{d+1}]|^2 \cdot p_{i_1} w_{i_1} p_{i_2} w_{i_2} \cdots p_{i_{d+1}} w_{i_{d+1}}.$$

González-Dávila, Dorta-Guerra and Ginebra (2007, Proposition 2.1) obtained essentially the same result. This can also be proved directly using the results from Rao (1973, Chapter 1). From Lemma 3.1 it is immediate that at least $(d+1)$ w_i's, as well as the corresponding p_i's, have to be positive for the determinant $f(\mathbf{p})$ to be nonzero. This implies if $\mathbf{p}$ is D-optimal, then $p_i < 1$ for each i. Theorem 3.1 below gives a sharper bound, $p_i \le \frac{1}{d+1}$ for each $i = 1, \ldots, 2^k$, for the optimal allocation. To check this we define for each $i = 1, \ldots, 2^k$,

$$f_i(z) = f\left(\frac{1-z}{1-p_i}p_1, \ldots, \frac{1-z}{1-p_i}p_{i-1}, z, \frac{1-z}{1-p_i}p_{i+1}, \ldots, \frac{1-z}{1-p_i}p_{2^k}\right), \quad 0 \le z \le 1. \quad (2)$$

Note that $f_i(z)$ is well defined for all $\mathbf{p}$ of interest. The next theorem characterizes all locally D-optimal allocations.

Theorem 3.1 *Suppose $f(\mathbf{p}) > 0$. Then $\mathbf{p}$ is D-optimal if and only if for each $i = 1, \ldots, 2^k$, one of the two conditions below is satisfied:*

(i) $p_i = 0$ *and* $f_i\left(\frac{1}{2}\right) \le \frac{d+2}{2^{d+1}} f(\mathbf{p})$;

(ii) $0 < p_i \le \frac{1}{d+1}$ *and* $f_i(0) = \frac{1-p_i(d+1)}{(1-p_i)^{d+1}} f(\mathbf{p})$.

A design that is minimally supported on $(d+1)$ points is called a saturated design. These designs are attractive because they require a minimum number of experimental settings. In our context, a design $\mathbf{p} = (p_1, \ldots, p_{2^k})'$ with a $2^k \times (d+1)$ design matrix X is called *saturated* if it contains exactly $(d+1)$ nonzero p_i's. Based on Lemma 3.1, a saturated design $\mathbf{p}$ with $p_{i_1} > 0, \ldots, p_{i_{d+1}} > 0$ is D-optimal if and only if $p_{i_1} = \cdots = p_{i_{d+1}} = \frac{1}{d+1}$. Yang, Mandal and Majumder (2011) found a necessary and sufficient condition for a saturated design to be D-optimal for 2^2 main-effects model. With the aid of Theorem 3.1, we provide a generalization for 2^k designs as follows. Note that $w_i > 0$ for each i for the commonly used link functions including logit, probit, and (complementary) log-log.

Theorem 3.2 *Assume $w_i > 0$, $i = 1, \ldots, 2^k$. Let $\mathbf{I} = \{i_1, \ldots, i_{d+1}\} \subset \{1, \ldots, 2^k\}$ be an index set satisfying $|X[i_1, \ldots, i_{d+1}]| \neq 0$. Then the saturated design satisfying $p_{i_1} = p_{i_2} = \cdots = p_{i_{d+1}} = \frac{1}{d+1}$ is D-optimal if and only if for each $i \notin \mathbf{I}$,*

$$\sum_{j \in \mathbf{I}} \frac{|X[\{i\} \cup \mathbf{I} \setminus \{j\}]|^2}{w_j} \le \frac{|X[i_1, i_2, \ldots, i_{d+1}]|^2}{w_i}.$$

For example, under the 2^2 main-effects model, $p_1 = p_2 = p_3 = 1/3$ is D-optimal if and only if $v_1 + v_2 + v_3 \leq v_4$, where $v_i = 1/w_i$, $i = 1, 2, 3, 4$. For the 2^3 main-effects model, $p_1 = p_4 = p_6 = p_7 = 1/4$ is D-optimal if and only if $v_1 + v_4 + v_6 + v_7 \leq 4 \min\{v_2, v_3, v_5, v_8\}$, and $p_2 = p_3 = p_5 = p_8 = 1/4$ is D-optimal if and only if $v_2 + v_3 + v_5 + v_8 \leq 4 \min\{v_1, v_4, v_6, v_7\}$.

Remark 1 In order to characterize the uniqueness of the optimal allocation, we define a matrix $X_w = [\mathbf{1},\ \mathbf{w} * \mathbf{1},\ \mathbf{w} * \gamma_2,\ \ldots,\ \mathbf{w} * \gamma_s]$, where $\mathbf{1}, \gamma_2, \ldots, \gamma_s$ are all distinct pairwise Schur products of the columns of the design matrix X, $\mathbf{w} = (w_1, \ldots, w_{2^k})'$, and "$*$" indicates Schur product. It can be seen that any two feasible allocations generate the same matrix $X'WX$ as long as their difference belongs to the null space of X_w. If $\text{rank}(X_w) < 2^k$, any criterion based on $X'WX$ yields not just a single solution but an affine set of solutions with dimension $2^k - \text{rank}(X_w)$. If $\text{rank}(X_w) = 2^k$, the D-optimal allocation $\mathbf{p}$ is unique. For example, for a 2^3 design the model consisting of all main effects and one two-factor interaction, or for a 2^4 design the model consisting of all main effects, all two-factor interactions, and one three-factor interaction, will lead to a unique D-optimal allocation.

3.2 EW D-optimal designs

In order to use the D-optimal designs, we need to specify initial values for the β_i's, which gives us the w_i's. In Section 5, we will examine the robustness of the D-optimal design to mis-specification of β_i's. In situations where reliable initial values are unavailable, if a prior distribution may be specified for the parameters, one may use Bayes optimal designs. An alternative to Bayes optimality is to replace the information matrix $X'WX$ in the criterion by its expectation $E(X'WX)$. This is one of several alternatives suggested by Atkinson, Donev and Tobias (2007). We call this EW optimality.

Definition: An *EW D-optimal* design is an optimal allocation of $\mathbf{p}$ that maximizes $|X'E(W)X|$.

One interpretation of EW optimality is that it is local optimality with w_i's replaced by their expectation. In this section we study the properties of EW-optimality.

A Bayes D-optimal design maximizes $E(\log|X'WX|)$ where the expectation is taken over the joint prior distribution of β_i's. Note that, by Jensen's inequality,

$$E\left(\log|X'WX|\right) \leq \log|X'E(W)X|$$

since $\log|X'WX|$ is concave in $\mathbf{w}$. Thus an EW D-optimal design maximizes an upper bound to the Bayesian D-optimality criterion.

We will show that EW-optimal designs are often almost as efficient as Bayes-optimal ones with respect to the Bayes optimality criterion, while realizing considerable savings in computation time. Furthermore, EW-optimal designs are highly robust in terms of maximum relative efficiencies (see Section 5 for details). Note that given link function g, $w_i = \nu(\eta_i) = \nu\left(\mathbf{x}_i'\boldsymbol{\beta}\right)$, $i = 1, \ldots, 2^k$, where $\nu = \left((g^{-1})'\right)^2 / \left[g^{-1}(1-g^{-1})\right]$, $\mathbf{x}_i$ is the ith row of the design matrix, and $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_d)'$.

Suppose the regression coefficients, $\beta_0, \beta_1, \ldots, \beta_d$ are independent, and $\beta_1, \ldots, \beta_d$ all have a symmetric distribution about 0 (not necessarily the same distribution), then it can be shown that the uniform design with $p_1 = \cdots = p_{2^k} = 2^{-k}$ is an EW D-optimal design for any given link function. While it is an interesting theoretical result that characterizes the EW D-optimal designs in terms of the regression coefficients, in most situations the experimenter will be able to specify the sign of the β_i's and possibly the range. If $\beta_i \in [0, \beta_{iu}]$ for each i, the uniform design will not be EW D-optimal in general, as illustrated in the following examples.

Example 3.1 Consider 2^2 experiment with main-effects model. Suppose the experimenter has the following prior information for the parameters: $\beta_0, \beta_1, \beta_2$ are independent, $\beta_0 \sim U[-1, 1]$, and $\beta_1, \beta_2 \sim U[0, 1]$. Under the logit link, $E(w_1) = E(w_4) = 0.187$, $E(w_2) = E(w_3) = 0.224$, and the EW D-optimal design is $\mathbf{p}_e = (0.239,\ 0.261,\ 0.261,\ 0.239)'$. The Bayes D-optimal design, which maximizes $\phi(\mathbf{p}) = E(\log|X'WX|)$, is $\mathbf{p}_o = (0.235,\ 0.265,\ 0.265,\ 0.235)'$. The relative efficiency of $\mathbf{p}_e$ with respect to $\mathbf{p}_o$ is

$$\exp\left\{\frac{\phi(\mathbf{p}_e) - \phi(\mathbf{p}_o)}{d+1}\right\} \times 100\% = \exp\left\{\frac{-4.80665 - (-4.80642)}{2+1}\right\} \times 100\% = 99.99\%.$$

The computational time cost for EW is 0.11 sec, while it is 5.45 secs for maximizing $\phi(\mathbf{p})$. It should also be noted that the relative efficiency of the uniform design $\mathbf{p}_u = (1/4, 1/4, 1/4, 1/4)'$ with respect to $\mathbf{p}_o$ is 99.88% for logit link, and is 89.6% for complementary log-log link (EW design's is 100.00%, see Remark 3).

Example 3.2 Consider 2^3 experiment with main-effects model. Suppose $\beta_0, \beta_1, \beta_2, \beta_3$ are independent, $\beta_0 \sim U[-3, 3]$, and $\beta_1, \beta_2, \beta_3 \sim U[0, 3]$. Then $E(w_1) = E(w_8) = 0.042$, $E(w_2) = E(w_3) = \cdots = E(w_7) = 0.119$. Under the logit link the EW D-optimal design is given by $\mathbf{p}_e = (0, 1/6, 1/6, 1/6,\ 1/6,\ 1/6,\ 1/6,\ 0)'$, and the Bayesian D-optimal design is $\mathbf{p}_o = (0.004, 0.165, 0.166, 0.165, 0.165, 0.166, 0.165, 0.004)'$. The relative efficiency of $\mathbf{p}_e$ with respect to $\mathbf{p}_o$ is 99.98%, while the relative efficiency of the uniform design is 94.39%. It is remarkable that it takes 2.39 seconds to find an EW solution while it takes 121.73 seconds to find $\mathbf{p}_o$. The difference in computational time is even more prominent for 2^4 case (24 seconds versus 3147 seconds).

3.3 Algorithms to search for locally D-optimal allocation

In this section, we develop efficient algorithms to search for locally D-optimal allocations with given w_i's. The same algorithms can be used for finding EW D-optimal designs.

3.3.1 Lift-one algorithm for maximizing $f(\mathbf{p}) = |X'WX|$

Here we propose the *lift-one* algorithm for searching locally D-optimal $\mathbf{p} = (p_1, \ldots, p_{2^k})'$ with given w_i's. The basic idea is that, for randomly chosen $i \in \{1, \ldots, 2^k\}$, we update p_i to p_i^* and all the other p_j's to $p_j^* = p_j \cdot \frac{1-p_i^*}{1-p_i}$. This technique is motivated by the coordinate descent algorithm (Zangwill, 1969). It is also in spirit similar to the idea of one-point correction in the literature (Wynn, 1970; Fedorov, 1972; Muller, 2007), where design points are added/adjusted one by one. The major advantage of the lift-one algorithm is that in order to determine an optimal p_i^*, we need to calculate $|X'WX|$ only once due to Lemma 3.1.

Lift-one algorithm:

1° Start with arbitrary $\mathbf{p}_0 = (p_1, \ldots, p_{2^k})'$ satisfying $0 < p_i < 1$, $i = 1, \ldots, 2^k$ and compute $f(\mathbf{p}_0)$.

2° Set up a random order of i going through $\{1, 2, \ldots, 2^k\}$.

3° For each i, determine $f_i(z)$ as in (A.8). In this step, either $f_i(0)$ or $f_i\left(\frac{1}{2}\right)$ needs to be calculated according to equation (2).

4° Define $\mathbf{p}_*^{(i)} = \left(\frac{1-z_*}{1-p_i}p_1, \ldots, \frac{1-z_*}{1-p_i}p_{i-1}, z_*, \frac{1-z_*}{1-p_i}p_{i+1}, \ldots, \frac{1-z_*}{1-p_i}p_{2^k}\right)'$, where z_* maximizes $f_i(z)$ with $0 \le z \le 1$ (see Lemma 7.2). Note that $f(\mathbf{p}_*^{(i)}) = f_i(z_*)$.

5° Replace $\mathbf{p}_0$ with $\mathbf{p}_*^{(i)}$, $f(\mathbf{p}_0)$ with $f(\mathbf{p}_*^{(i)})$.

6° Repeat 2° $\sim$ 5° until convergence, that is, $f(\mathbf{p}_0) = f(\mathbf{p}_*^{(i)})$ for each i.

While in all examples that we studied, the lift-one algorithm converges very fast, we do not have a proof of convergence. There is a modified lift-one algorithm, which is only slightly slower, that can be shown to converge. This algorithm can be described as follows. For the $10m$th iteration and a fixed order of $i = 1, \ldots, 2^k$ we repeat steps 3° $\sim$ 5°, $m = 1, 2, \ldots$, if $\mathbf{p}_*^{(i)}$ is a better allocation found by the lift-one algorithm than the allocation $\mathbf{p}_0$, instead of updating $\mathbf{p}_0$ to $\mathbf{p}_*^{(i)}$ immediately, we obtain $\mathbf{p}_*^{(i)}$ for each i, and replace $\mathbf{p}_0$ with the first best one among $\left\{\mathbf{p}_*^{(i)}, i = 1, \ldots, 2^k\right\}$. For iterations other than the $10m$th, we follow the original lift-one algorithm update.

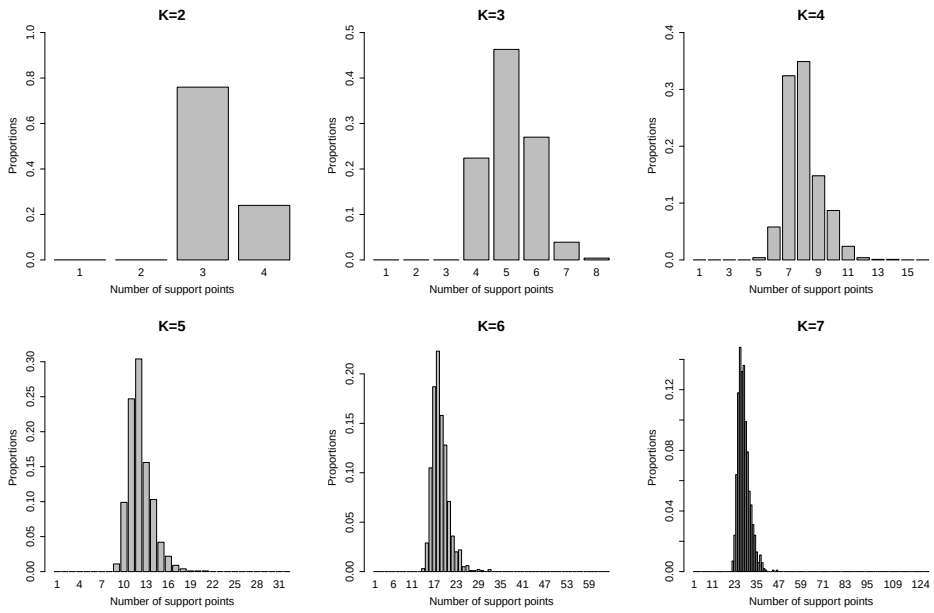

Figure 1: Number of support points in the optimal design (based on 1000 simulations)

Theorem 3.3 *When the lift-one algorithm or the modified lift-one algorithm converges, the converged allocation* $\mathbf{p}$ *maximizes* $|X'WX|$ *on the set of feasible allocations. Furthermore, the modified lift-one algorithm is guaranteed to converge.*

Our simulation studies indicate that as k grows, the optimal designs produced by our lift-one algorithm is supported only on a fraction of all the 2^k design points. To illustrate this, we randomly generated the regression coefficients iid from $U(-3, 3)$ and applied our algorithm to find the optimal designs under logit link. Figure 1 gives histograms of number of support points in the optimal designs. For example, with $k = 2$, 76% of the designs are supported on three points only and 24% of them are supported on all four points. As k becomes larger, the number of support points moves towards a smaller fraction.

As demonstrated in Table 1, our lift-one algorithm is much faster than commonly used optimization techniques including Nelder-Mead, quasi-Newton, conjugate-gradient, and simulated annealing (for a comprehensive reference, see Nocedal and Wright (1999)). Our algorithm finds the optimal solution with more accuracy as well. For example, for a 2^4 design problem, the relative efficiency of the design found by the Nelder-Mead method with respect to the design found by our algorithm is only 95% on average. There are 10% of the cases with relative efficiencies less than 75%. When the number of factors becomes large, our investigation indicates that some algorithms do not converge within a reasonable amount of time or are not applicable due to unavailability of the gradient of the objective function. These are denoted by "NA" in Table 1.

Table 1: Performance of the lift-one algorithm (CPU time in seconds for 100 simulated $\mathbf{w}$)

	Algorithms				
	Nelder-Mead	quasi-Newton	conjugate gradient	simulated annealing	lift-one
2^2 Designs	0.77	0.41	4.68	46.53	0.17
2^3 Designs	46.75	63.18	925.5	1495	0.46
2^4 Designs	125.9	NA	NA	NA	1.45

3.3.2 Algorithm for maximizing $|X'WX|$ with integer solutions

To maximize $|X'WX|$, an alternative algorithm, called *exchange algorithm* (see Supplementary Materials for detailed description), is to adjust p_i and p_j simultaneously for randomly chosen index pair (i, j). The original idea of exchange was suggested by Fedorov (1972). Due to Lemma 3.1, the optimal adjusted (p_i^*, p_j^*) can be obtained easily by maximizing a quadratic function. Unlike the lift-one algorithm, the exchange algorithm can be applied to search for integer-valued optimal allocation $\mathbf{n} = (n_1, \ldots, n_{2^k})'$, where $\sum_i n_i = n$.

Exchange algorithm for integer-valued allocations:

1° Start with initial design $\mathbf{n} = (n_1, \ldots, n_{2^k})'$ such that $f(\mathbf{n}) > 0$.

2° Set up a random order of (i, j) going through all pairs

$$\{(1,2), (1,3), \ldots, (1,2^k), (2,3), \ldots, (2^k - 1, 2^k)\}$$

3° For each (i, j), let $m = n_i + n_j$. If $m = 0$, let $\mathbf{n}_{ij}^* = \mathbf{n}$. Otherwise, calculate $f_{ij}(z)$ as given in equation (A.11). Then let

$$\mathbf{n}_{ij}^* = (n_1, \ldots, n_{i-1}, z_*, n_{i+1}, \ldots, n_{j-1}, m - z_*, n_{j+1}, \ldots, n_{2^k})$$

where the integer z_* maximizes $f_{ij}(z)$ with $0 \le z \le m$ according to Lemma 7.4 in the Appendix. Note that $f(\mathbf{n}_{ij}^*) = f_{ij}(z_*) \ge f(\mathbf{n}) > 0$.

4° Repeat 2° $\sim$ 3° until convergence (no more increase in terms of $f(\mathbf{n})$ by any pairwise adjustment).

As expected, the integer-valued optimal allocation $(n_1, \ldots, n_{2^k})'$ is consistent with the proportion-valued allocation $(p_1, \ldots, p_{2^k})'$ for large n. For small n, the result may be used for fractional design problem in Section 4. It should be noted that the exchange algorithm with slight modification is guaranteed to converge to the optimal allocation when searching for proportions but not for integer-valued solutions, especially when n is small compared to 2^k.

In terms of finding optimal proportions, the exchange algorithm produces essentially the same results as the lift-one algorithm, although the former is relatively slower. For examples, based on 1000 simulations, the ratio of computational time of the exchange algorithm over the lift-one algorithm is 6.2, 10.2, 16.8, 28.8, 39.5 and 51.3 for $k = 2, \ldots, 7$ respectively. Note that it requires 2.02, 5.38, 19.2, 84.3, 352, and 1245 seconds respectively to finish the 1000 simulations using the lift-one algorithm on a regular PC with 2.26GHz CPU and 2.0G memory. As the total number of factors k becomes larger, the computation is more intensive. Detailed study of the computational properties of the proposed algorithms is a topic of future research.

4 Fractional Factorial Designs

Unless the number of factors k is very small, the total number of experimental conditions 2^k is large, and it could be impossible to take observations at all 2^k level combinations. The reason is twofold: the total number of observations n would become large, as well as the number of level changes which are expensive in many applications. So when k is large, the experimenter may have to consider fractional designs. For linear models, the accepted practice is to use regular fractions due to the many desirable properties like minimum aberration and optimality. We will show that in our setup the regular fractions are not always optimal. First, we will examine the situations when they are optimal.

We use 2^3 designs for illustration. More discussion on general cases can be found in Section 5.1. The design matrix for 2^3 main-effects model consists of the first four columns of X given in equation (1) and w_j represents the information corresponding to the jth row of that matrix. The maximum number of experimental conditions is fixed at a number less than 8, and the problem is to identify the experimental conditions (rows of X) and corresponding p_i's that optimize the design optimality criterion. Half fractions use 4 experimental conditions (hence the design is uniform). The fractions defined by rows $\{1, 4, 6, 7\}$ and $\{2, 3, 5, 8\}$ are regular fractions, given by the defining relations $1 = ABC$ and $-1 = ABC$ respectively. If the initial values of all the regression coefficients corresponding to the design factors are zeros, it leads to the case where all the w_i's are equal and effectively reduces to the linear model, where the regular fractions are D-optimal. The following Theorem identifies the necessary and sufficient conditions for regular fractions to be D-optimal in terms of w_i's and are valid for all link functions.

Theorem 4.1 *For the 2^3 main-effects model, suppose $\beta_1 = 0$. This implies $w_1 = w_5$, $w_2 = w_6$, $w_3 = w_7$, and $w_4 = w_8$. The regular fractions $\{1, 4, 6, 7\}$ and $\{2, 3, 5, 8\}$ are D-optimal half-fractions if and only if*

$$4 \ \min\{w_1, w_2, w_3, w_4\} \geq \max\{w_1, w_2, w_3, w_4\}.$$

Further suppose $\beta_2 = 0$. Then $w_1 = w_3 = w_5 = w_7$ and $w_2 = w_4 = w_6 = w_8$. The

two regular half-fractions mentioned above are D-optimal half-fractions if and only if $4\min\{w_1, w_2\} \geq \max\{w_1, w_2\}$.

Example 4.1 Under logit link, consider the 2^3 main-effects model with $\beta_1 = \beta_2 = 0$. The regular half-fractions $\{1, 4, 6, 7\}$ and $\{2, 3, 5, 8\}$ are D-optimal half-fractions if and only if one of the following happens:

$$\begin{aligned} &(i) \quad |\beta_3| \leq \log 2 \\ &(ii) \quad |\beta_3| > \log 2 \quad \text{and} \quad |\beta_0| \leq \log\left(\frac{2e^{|\beta_3|} - 1}{e^{|\beta_3|} - 2}\right). \end{aligned} \tag{3}$$

When the regular half-fractions are not optimal, the goal is to find $\{i_1, i_2, i_3, i_4\}$ that maximizes $|X[i_1, i_2, i_3, i_4]|^2 w_{i_1} w_{i_2} w_{i_3} w_{i_4}$ based on Lemma 3.1. Recall that in this case there are only two distinct w_i's. If $\beta_0\beta_3 > 0$, w_i's corresponding to $\{2, 4, 6, 8\}$ are larger than others, so this fraction given by $C = -1$ will maximize $w_{i_1} w_{i_2} w_{i_3} w_{i_4}$ but this leads to a singular design matrix. It is not surprising that the D-optimal half-fractions are "close" to the design $\{2, 4, 6, 8\}$, and are in fact given by the 16 designs each consisting of three elements from $\{2, 4, 6, 8\}$ and one from $\{1, 3, 5, 7\}$. For $\beta_0\beta_3 < 0$, D-optimal half-fractions are similarly obtained from the fraction $C = +1$.

Figure 2 below partitions the parameter space for 2^3 main-effects logit model. The left panel corresponds to the case $\beta_1 = \beta_2 = 0$. Here the parameters in the middle region would make the regular fractions D-optimal whereas the top-right and bottom-left regions correspond to the case $\beta_0\beta_3 > 0$. Similarly the other two regions correspond to the case $\beta_0\beta_3 < 0$. The right panel of Figure 2 is for the case $\beta_1 = 0$ and shows the contour plots for the largest $|\beta_0|$'s that would make the regular fractions D-optimal. (For details, see the Supplementary Materials of this paper.) Along with Figure 2, conditions (3) and (S.10) in Supplementary Materials indicate that if β_1,β_2 and β_3 are small then the regular fractions are preferred (see also Table 2 below). However, when at least one $|\beta_i|$ is large, the information is not uniformly distributed over the design points and the regular fractions may not be optimal.

In general, when all the β_i's are nonzero, the regular fractions given by the rows $\{1, 4, 6, 7\}$ or $\{2, 3, 5, 8\}$ are not necessarily the optimal ones. To illustrate this, we simulate the regression coefficients β_0, β_1, β_2, β_3 independently from different distributions and calculate the corresponding $\mathbf{w}$'s under logit, probit and complementary log-log links for 10,000 times each. For each $\mathbf{w}$, we find the best design supported on 4 distinct rows of the design matrix. By Lemma 3.1, any such design has to be uniform in order to be D-optimal. Table 2 reports the percentages of times each of those designs turn out to be the optimal ones for the logit model. The results are somewhat similar for the other links. It shows that the regular fractions are optimal when the β_i's are close to zero. In Table 2, we only reported the fractions which turned out to be D-optimal for more than 15% of the times with the exception of regular fractions. For the 2^4 case, the results are similar, that is, when the β_i's are

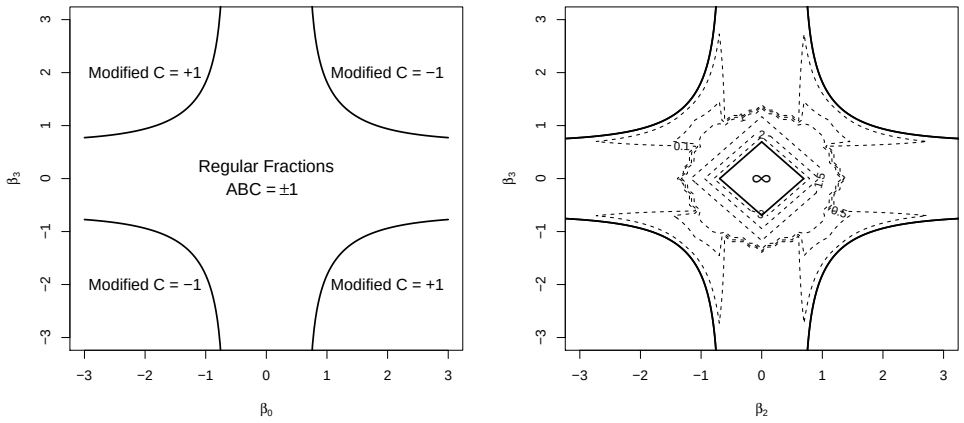

Figure 2: Partitioning of the parameter space

Table 2: Distribution of D-optimal half-fractions under 2^3 main-effects model

Rows	Percentages					
	$\beta_0 \sim$		$U(-10,10)$			$N(0,5)$
Simulation	$\beta_1 \sim$	$U(-.3,.3)$	$U(-3,3)$	$U(-3,0)$	$U(0,1)$	$N(1,1)$
Setup	$\beta_2 \sim$	$U(-.3,.3)$	$U(0,3)$	$U(0,3)$	$U(0,3)$	$N(2,1)$
	$\beta_3 \sim$	$U(-.3,.3)$	$U(1,5)$	$U(-2,2)$	$U(0,5)$	$N(3,1)$
1467		42.65 (99.5)	0.07 (0.22)	0.86 (2.18)	0.95 (2.61)	0.04 (2.89)
2358		42.02 (99.5)	0.04 (0.23)	0.68 (2.12)	1.04 (2.56)	0.08 (2.86)
1235			16.78		35.62	21.50
1347				19.98		
1567			17.45	19.21		
2348			17.54	19.11		
2568				20.01		
4678			16.12		35.41	21.65

nonzeros, the performance of the regular fractions given by $1 = \pm ABCD$ are not very efficient in general.

Relative efficiency of a design is defined as $1 - R(\mathbf{p}, \mathbf{w})$ where $R(\mathbf{p}, \mathbf{w})$ is the relative loss of efficiency given in equation (4) of Section 5.1. In Table 2, we have provided, within parenthesis, the percentages of times the regular fractions were at least 95% efficient compared to the best half-fractions. It is clear that when the regular fractions are not D-optimal, they are usually not highly efficient either. On the other hand, for each of the five situations described in Table 2, we also calculated the corresponding EW D-optimal half-fractions. For all of the five cases, including the highly asymmetric fifth scenario, the condition for Theorem 4.1 is satisfied and the regular fractions are the EW D-optimal half-fractions.

Remark 2 Consider the problem of obtaining the locally D-optimal fractional factorial designs when the number of experimental settings (m, say) is fixed. If the total

number of factors under consideration is not too large, one can always calculate the D-efficiencies of all fractions and choose the best one. However, we need a better strategy for even moderately large number of factors. One such strategy would be to choose the m largest w_i's and the corresponding rows, since the w_i's are the information at that point. Another one could be to use our algorithms discussed in Section 3.3 for finding an optimal allocation for the full factorial designs first. Then choose the m largest p_i's and scale them appropriately. One has to be careful in order to avoid designs which would not allow the estimation of the model parameters. In this case, the exchange algorithm described in Section 3.3.2 may be used to choose the fraction with given m experimental units. Our simulations (not presented here) shows that both of these methods perform satisfactorily with the second method giving designs which are generally more than 95% efficient for four factors with main-effects model. This method will be used for computations in the next section.

5 Robustness

In this section, we will check the robustness of locally D-optimal designs over the initial parameter values, for both full factorial and fractional factorial setups.

5.1 Most robust saturated designs

Saturated designs have been studied extensively. For continuous, quantitative factors, these designs are D-optimal for many linear and non-linear models. In our setup of qualitative factors, saturated designs are attractive since they use the minimal number, $(d+1)$, of experimental conditions. In many applications, level changes are expensive, hence fewer experimental conditions are desirable. In this section, we will examine the robustness of saturated designs. Since here the design matrix has $(d+1)$ columns, it is straightforward to find the locally D-optimal saturated designs as given in Theorem 5.1, which immediately follows from Lemma 3.1.

Theorem 5.1 *Let* $\mathbf{I} = \{i_1, \ldots, i_{d+1}\} \subset \{1, \ldots, 2^k\}$ *be an index set. A design* $\mathbf{p}_I = (p_1, \ldots, p_{2^k})'$ *satisfying* $p_i = 0, \forall i \notin I$ *is a D-optimal saturated design if and only if*

$$p_{i_1} = \cdots = p_{i_{d+1}} = \frac{1}{d+1} \text{ and } \mathbf{I} \text{ maximizes } |X[i_1, \ldots, i_{d+1}]|^2 w_{i_1} \cdots w_{i_{d+1}}.$$

For investigating the robustness of these designs, let us denote the D-criterion value as $\psi(\mathbf{p}, \mathbf{w}) = |X'WX|$ for given $\mathbf{w} = (w_1, \ldots, w_{2^k})'$ and $\mathbf{p} = (p_1, \ldots, p_{2^k})'$. Suppose $\mathbf{p}_w$ is a D-optimal allocation with respect to $\mathbf{w}$. Then the relative loss of efficiency of $\mathbf{p}$ with respect to $\mathbf{w}$ can be defined as

$$R(\mathbf{p}, \mathbf{w}) \;=\; 1 - \left(\frac{\psi(\mathbf{p}, \mathbf{w})}{\psi(\mathbf{p}_w, \mathbf{w})} \right)^{\frac{1}{d+1}}. \tag{4}$$

Let us define the maximum relative loss of efficiency of a given design $\mathbf{p}$ with respect to a specified region $\mathcal{W}$ of $\mathbf{w}$ by

$$R_{\max}(\mathbf{p}) = \max_{\mathbf{w}\in\mathcal{W}} R(\mathbf{p}, \mathbf{w}). \tag{5}$$

It can be shown that $\mathcal{W}$ takes the form $[a,\ b]^{2^k}$ for 2^k main-effects model if the range of each of the regression coefficients is an interval symmetric about 0. For example, for a 2^4 main-effects model, if all the regression coefficients range between $(-3, 3)$, then $\mathcal{W} = [3.06 \times 10^{-7},\ 0.25]^{16}$ for logit link, $[8.33 \times 10^{-49}, 0.637]^{16}$ for probit link, or $(0, 0.648]^{16}$ for (complementary) log-log link. This justifies the choice of the range of w_i's in Theorem 5.2 below. A design which minimizes the maximum relative loss of efficiency will be called *most robust*.

Theorem 5.2 *Suppose $k \geq 3$ and $d(d+1) \leq 2^{k+1} - 4$. Suppose $w_i \in [a,\ b]$, $i = 1, \ldots, 2^k$, $0 < a < b$. Let $\mathbf{I} = \{i_1, \ldots, i_{d+1}\}$ be an index set which maximizes $|X[i_1, i_2, \ldots, i_{d+1}]|^2$. Then the design $\mathbf{p}_I = (p_1, \ldots, p_{2^k})'$ satisfying $p_{i_1} = \cdots = p_{i_{d+1}} = \frac{1}{d+1}$ is a most robust saturated design with maximum relative loss of efficiency $1 - \frac{a}{b}$.*

Based on Theorem 5.2, the maximum relative loss of efficiency depends on the possible range of w_i's. The result is practically meaningful only if the interval $[a, b]$ is bounded away from 0. Figure 3 provides some idea about the possible bounds of w_i's for commonly used link functions. For example, for 2^3 designs with main-effects model, suppose we know that $0.105 \leq w_i \leq 0.25$ under logit link (see Remark 4.1.1 of Yang, Mandal and Majumdar (2012)), then the maximum relative loss of efficiency of the regular half-fractional design satisfying $p_1 = p_4 = p_6 = p_7 = 1/4$ is $1 - 0.105/0.25 = 58\%$. The more certain we are about the range of w_i's, the more useful the result will be.

Note that for main-effects models, the condition $d(d+1) \leq 2^{k+1} - 4$ in Theorem 5.2 is guaranteed by $k \geq 3$. A most robust saturated design can be obtained by searching an index set $\{i_1, \ldots, i_{d+1}\}$ which maximizes $|X[i_1, i_2, \ldots, i_{d+1}]|^2$. Note that such an index set is usually not unique. Based on Lemma 7.3, if the index set $\{i_1, \ldots, i_{d+1}\}$ maximizes $|X[i_1, \ldots, i_{d+1}|^2$, there always exists another index set $\{i'_1, \ldots, i'_{d+1}\}$ such that $|X[i_1, \ldots, i_{d+1}]|^2 = |X[i'_1, \ldots, i'_{d+1}]|^2$. It should also be noted that a most robust saturated design may involve a set of experimental conditions $\{i_1, \ldots, i_{d+1}\}$ which does not maximize $|X[i_1, \ldots, i_{d+1}]|^2$. For example, consider a 2^{3-1} design with main-effects model. Suppose $w_i \in [a, b]$, $i = 1, \ldots, 8$. If $4a > b$, then the most robust saturated designs are the regular 2^{3-1} fractional ones. Otherwise, if $4a \leq b$, then any uniform design restricted to $\{i_1, i_2, i_3, i_4\}$ satisfying $|X[i_1, i_2, i_3, i_4]| \neq 0$ is a most robust saturated one.

5.2 Robustness of uniform designs

A design is called "uniform" if the allocation of experimental units is the same for all points in the support of the design. Yang, Mandal, and Majumdar (2011) showed

Table 3: Relative loss of efficiency of 2^4 uniform design

	Percentages											
	$\beta_0 \sim U(-3,3)$			$U(-1,1)$			$U(-3,0)$			$N(0,5)$		
	$\beta_1 \sim U(-1,1)$			$U(0,1)$			$U(1,3)$			$N(0,1)$		
Simulation	$\beta_2 \sim U(-1,1)$			$U(0,1)$			$U(1,3)$			$N(2,1)$		
Setup	$\beta_3 \sim U(-1,1)$			$U(0,1)$			$U(-3,-1)$			$N(-.5,2)$		
	$\beta_4 \sim U(-1,1)$			$U(0,1)$			$U(-3,-1)$			$N(-.5,2)$		
Quantiles	(I)	(II)	(III)	(I)	(II)	(III)	(I)	(II)	(III)	(I)	(II)	(III)
R_{99}	.348	.353	.348	.146	.111	.112	.503	.273	.299	.650	.864	.726
R_{95}	.299	.304	.299	.128	.094	.093	.495	.251	.256	.617	.788	.670
R_{90}	.271	.274	.271	.117	.084	.085	.488	.239	.233	.589	.739	.629

Note: (I) $= R_{100\alpha}(\mathbf{p}_u)$, (II) $= \min_{1\le s\le 1000} R_{100\alpha}(\mathbf{p}_s)$, (III) $= R_{100\alpha}(\mathbf{p}_e)$.

that for a 2^2 main-effects model, the uniform design is the most robust design in terms of maximum relative loss of efficiency. In this section, we use simulation studies to examine the robustness of uniform designs and EW D-optimal designs for higher order cases.

For illustration, we use a 2^4 main-effects model. We simulate $\beta_0, \ldots, \beta_4$ from different distributions 1000 times each and calculate the corresponding $\mathbf{w}$'s, denoted by $\mathbf{w}_1$, ..., $\mathbf{w}_{1000}$. For each $\mathbf{w}_s$, we use the algorithm described in Section 3.3.1 to obtain a D-optimal allocation $\mathbf{p}_s$. For any allocation $\mathbf{p}$, let $R_{100\alpha}(\mathbf{p})$ denote the αth quantile of the set of relative loss of efficiencies $\{R(\mathbf{p}, \mathbf{w}_s),\ s = 1, \ldots, 1000\}$. Thus $R_{100}(\mathbf{p}) = R_{\max}(\mathbf{p})$ which is the $R_{\max}$ defined in (5) with $\mathcal{W} = \{\mathbf{w}_1, \ldots, \mathbf{w}_{1000}\}$. Since $\mathcal{W}$ here is simulated, the quantities $R_{99}(\mathbf{p})$ or $R_{95}(\mathbf{p})$ are more reliable in measuring the robustness of $\mathbf{p}$.

Table 3 compares the $R_{100\alpha}$ of the uniform design $\mathbf{p}_u = (1/16, \ldots, 1/16)'$ with the minimum of $R_{100\alpha}(\mathbf{p}_s)$ for $s = 1, \ldots, 1000$, as well as the $R_{100\alpha}$ of the EW design $\mathbf{p}_e$. In this table, if the values of column (I) is smaller than those of column (II), then we can conclude that the uniform design is better than all the D-optimal designs in terms of the quantiles of relative loss of efficiency, which happens in many situations. It is a strong evidence supporting that the uniform design $\mathbf{p}_u$ is one of the most robust ones if the β_i's occupy symmetric range around zero. This is consistent with the conclusion of Cox (1988).

However, there are situations where the uniform design does not perform well, as illustrated by the two middle blocks of Table 3. If the signs of the regression coefficients are known, it is advisable not to use the uniform design. For many practical applications, the experimenter will have some idea about the direction of effects of factors, which in statistical terms determines the signs of the regression coefficients. For these situations, it turns out that the performance of the EW D-optimal designs is comparable to that of the most robust designs, even when the uniform design does not perform well (see columns (III) in Table 3, where $\mathbf{p}_e$ is the EW design). Hence we recommend the use of EW D-optimal designs when the experimenter has some idea

about the signs of β_i's. Uniform designs are recommended in the absence of prior knowledge of the regression parameters.

Now consider the uniform designs restricted to regular fractions. Again we use 2^4 main-effects model as illustration and consider the uniform designs restricted to the regular half-fractions identified by $1 = \pm ABCD$. We performed simulations as above and our conclusions are similar, that is, uniform designs on regular fractions are among the most robust ones if the regression parameters are supported symmetrically around zero but they may not perform well if the β_i's are either positive or negative.

6 Examples and Discussion

In this section we make some remarks, with illustrations, about our results and their use.

Remark 3 The logit link is the most commonly used link in practice. The situation under this link function is close to that in the linear model case because w_i's do not vary much ($0 < w_i \leq 0.25$), hence the uniform design is expected to perform not too poorly, since they are optimal for the linear model case. But this is not the case for other link functions. In general, the performance of the logit and probit links are similar, while that of the complementary log-log link is somewhat different from others. For example, if we repeat Example 3.1 but under complementary log-log link, the relative efficiency of the uniform design is only 89.6% with respect to Bayes optimal design. Figure 3 provides a graphical display of the function ν for commonly used link functions. As seen from the figure, complementary log-log link function is not symmetric about 0. This explains the poor performance of the uniform design under this link. Nevertheless, the EW D-optimal designs are still highly efficient across different link functions. Repeating Example 3.1 with commonly used link functions, the relative efficiencies of EW designs with respect to the corresponding Bayesian optimal designs are 99.99% (logit link), 99.94% (probit link), 99.77% (log-log link), and 100.00% (complementary log-log link), respectively.

Example 6.1 Consider the Windshield molding experiment discussed in Section 1. By analyzing the data presented in Hamada and Nelder (1997), we get an estimate of the unknown parameter as $\hat{\boldsymbol{\beta}} = (1.77, -1.57, 0.13, -0.80, -0.14)'$ under logit link. If one wants to conduct a follow-up experiment on half-fractions, then it is sensible to use the knowledge obtained by analyzing the data. In this case, the efficiency of the original 2^{4-1}_{III} design is 78% of the locally D-optimal design if $\hat{\boldsymbol{\beta}}$ were the true value of the unknown parameter $\boldsymbol{\beta}$. Instead of taking $\hat{\boldsymbol{\beta}}$ exactly if a point close to it is taken as the initial value, say, $\boldsymbol{\beta} = (2, -1.5, 0.1, -1, -0.1)'$, the D-optimal design $\mathbf{p}_a$ is given in Table 4. A reasonable option is to consider a range for the possible values of the regression parameters, namely, $(1, 3)$ for β_0, $(-3, -1)$ for β_1, $(-0.5, 0.5)$ for β_2, β_4, and

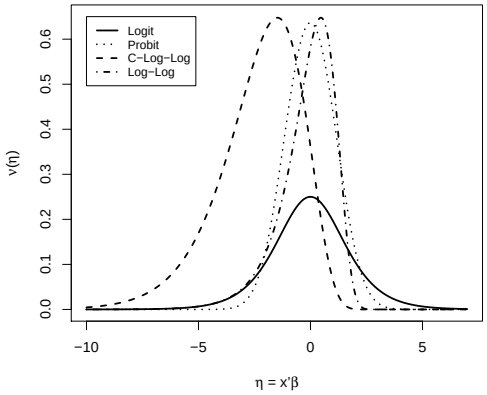

Figure 3: $w_i = \nu(\eta_i) = \nu(\mathbf{x}_i'\boldsymbol{\beta})$ for commonly used link functions

$(-1, 0)$ for β_3. For this choice of range for the parameter values with independence and uniform distributions, the EW D-optimal half-fractional design $\mathbf{p}_e$ is given in Table 4 too. Note that if we regard $\hat{\boldsymbol{\beta}}$ as the true value of the parameter, the relative efficiency of $\mathbf{p}_a$ increases to 99%, whereas that of $\mathbf{p}_e$ increases to 98%. We have also calculated the linear predictor η and success probability π for all possible experimental settings. It seems that a good fraction would not favor high success probabilities very much. This is one of the main differences between the design reported by Hamada and Nelder (denoted by $\mathbf{p}_{HN}$) and our designs (denoted by $\mathbf{p}_a$ and $\mathbf{p}_e$). Note that these two designs have six rows in common.

Example 6.2 Consider the "Project Quit" experiment discussed in Section 1. The authors noted that a full factorial experiment with 6 factors requiring 64 different settings was not feasible. So they conducted a 2^{6-2}_{IV} experiment defined by $E = ABC$ and $F = ACD$. The authors provided us with a dataset which was very similar to that one presented in their paper. By analyzing the data, we get an estimate of the unknown parameter as $\hat{\boldsymbol{\beta}} = (-1.10, 0.01, 0.06, 0.09, 0.02, 0.10, -0.04)'$ under logit link. Unlike the previous example, here the efficiency of the original design used by Nair *et al.* (2008) is more than 99% of the locally D-optimal design if $\hat{\boldsymbol{\beta}}$ were the true value of the unknown parameter $\boldsymbol{\beta}$. Hence, if someone wants to conduct a future experiment under similar set up, it will be advisable to use the same design. In this case, only factors C and E come out to be significant and that information can be used to reduce the number of factors, if needed.

Example 6.3 Consider the Mice experiment discussed in Section 1. The design used by Seo, Goldschmidt-Clermont and West (2007) is very efficient if all the regression coefficients of the factors are close to zeros, which leads to equal w_i's and corresponds to a linear model. However, their design is not very efficient if the β_i's are not close to zero. To illustrate this, we simulate 1000 $\boldsymbol{\beta}$'s from different distributions, and for each of them calculate the optimal allocation. Then we compare the efficiency of each

Table 4: Optimal half-fraction design for Windshield Molding Experiment

Row	A	B	C	D	η	π	$\mathbf{p}_{HN}$	$\mathbf{p}_a$	$\mathbf{p}_e$
5	+1	−1	+1	+1	-0.87	0.295		0.044	0.184
1	+1	+1	+1	+1	-0.61	0.352	0.125	0.178	0.011
6	+1	−1	+1	−1	-0.59	0.357	0.125	0.178	0.011
2	+1	+1	+1	−1	-0.33	0.418		0.059	0.184
7	+1	−1	−1	+1	0.73	0.675	0.125	0.163	
3	+1	+1	−1	+1	0.99	0.729			0.195
8	+1	−1	−1	−1	1.01	0.733			0.195
4	+1	+1	−1	−1	1.27	0.781	0.125	0.147	
13	−1	−1	+1	+1	2.27	0.906	0.125	0.158	0.111
9	−1	+1	+1	+1	2.53	0.926			
14	−1	−1	+1	−1	2.55	0.928			
10	−1	+1	+1	−1	2.81	0.943	0.125	0.074	0.110
15	−1	−1	−1	+1	3.87	0.980			
11	−1	+1	−1	+1	4.13	0.984	0.125		
16	−1	−1	−1	−1	4.15	0.984	0.125		
12	−1	+1	−1	−1	4.41	0.988			

Table 5: Performance of Seo et al.'s design

	Relative Efficiencies				
	$\beta_0 \sim$	$U(-10,10)$			
	$\beta_1 \sim$	$U(-.3,.3)$	$U(-3,3)$	$U(-3,3)$	$U(-3,0)$	$U(0,1)$
Simulation	$\beta_2 \sim$	$U(-.3,.3)$	$U(-3,3)$	$U(0,3)$	$U(0,3)$	$U(0,2)$
Setup	$\beta_3 \sim$	$U(-.3,.3)$	$U(-3,3)$	$U(1,3)$	$U(-2,2)$	$U(0,3)$
	$\beta_4 \sim$	$U(-.3,.3)$	$U(-3,3)$	$U(2,5)$	$U(-1,1)$	$U(0,5)$
Mean		0.940	0.503	0.446	0.566	0.547
SD		0.029	0.121	0.090	0.121	0.118

of these optimal designs with that of the design used by Seo et al. (2007) and report the mean and standard deviations of those efficiencies in Table 5.

As noted in the introduction and justified by the results in Section 5.2, if the experimenter has no prior knowledge of the parameters then it is recommended to use the uniform design. Otherwise, EW optimal designs should be used because they are very robust in terms of minimizing the maximum relative loss. In the EW optimality, we replace the w_i's by their expectations. Note that taking the average of w_i's is not same as taking the average of β_i's. Let us illustrate it with 2^4 design with main-effects model. Table 6 below uses the notations from Table 3. Suppose $\beta_0 \sim U(-3,0)$, $\beta_1, \beta_2 \sim U(1,3)$, $\beta_3, \beta_4 \sim U(-3,-1)$, and the β_i's are independent. It is clear that uniform design performs much worse compared to the most robust design, while the performance of the EW D-optimal design is comparable with the best design. The last column corresponds to the locally D-optimal design where the initial value of the parameter is taken to be the midpoints of the ranges of β_i's mentioned above. Clearly

Table 6: Performance of different designs for 2^4 main-effects model

	Uniform	Most robust	EW	$E(\beta)$
R_{99}	0.503	0.273	0.299	0.331
R_{95}	0.495	0.251	0.256	0.284
R_{90}	0.488	0.239	0.233	0.251

this is worse than the EW optimal design.

Acknowledgment

The research was in part supported by NSF Grant DMS-09-05731. The authors thank Bibhas Chakraborty for providing the data from Project Quit.

Appendix

We need two lemmas before the proof of Theorem 3.1.

Lemma 7.1 *Suppose* $\mathbf{p} = (p_1, \ldots, p_{2^k})'$ *satisfies* $f(\mathbf{p}) > 0$. *Given* $i = 1, \ldots, 2^k$,

$$f_i(z) = az(1-z)^d + b(1-z)^{d+1}, \tag{A.8}$$

for some constants a *and* b. *If* $p_i > 0$, $b = f_i(0)$, $a = \frac{f(\mathbf{p}) - b(1-p_i)^{d+1}}{p_i(1-p_i)^d}$; *otherwise,* $b = f(\mathbf{p})$, $a = f_i\left(\frac{1}{2}\right) \cdot 2^{d+1} - b$. *Note that* $a \geq 0$, $b \geq 0$, *and* $a + b > 0$. □

Lemma 7.2 *Let* $h(z) = az(1-z)^d + b(1-z)^{d+1}$ *with* $0 \leq z \leq 1$ *and* $a \geq 0, b \geq 0, a+b > 0$. *If* $a > b(d+1)$, *then* $\max_z h(z) = \left(\frac{d}{a-b}\right)^d \left(\frac{a}{d+1}\right)^{d+1}$ *at* $z = \frac{a-b(d+1)}{(a-b)(d+1)} < 1$. *Otherwise,* $\max_z h(z) = b$ *at* $z = 0$. □

Proof of Theorem 3.1: Note that $f(\mathbf{p}) > 0$ implies and $0 \leq p_i < 1$ for each $i = 1, \ldots, 2^k$. Since $\sum_i p_i = 1$, without any loss of generality, we assume $p_{2^k} > 0$. Define $\mathbf{p}_r = (p_1, \ldots, p_{2^k-1})'$, and $f^{(r)}(\mathbf{p}_r) = f\left(p_1, \ldots, p_{2^k-1}, 1 - \sum_{i=1}^{2^k-1} p_i\right)$.

For $i = 1, \ldots, 2^k - 1$, let $\boldsymbol{\delta}_i^{(r)} = (-p_1, \ldots, -p_{i-1}, 1-p_i, -p_{i+1}, \ldots, -p_{2^k-1})'$. Then $f_i(z) = f^{(r)}(\mathbf{p}_r + u\boldsymbol{\delta}_i^{(r)})$ with $u = \frac{z-p_i}{1-p_i}$. Since the determinant $|(\boldsymbol{\delta}_1^{(r)}, \ldots, \boldsymbol{\delta}_{2^k-1}^{(r)})| = p_{2^k} \neq 0$, $\boldsymbol{\delta}_1^{(r)}, \ldots, \boldsymbol{\delta}_{2^k-1}^{(r)}$ are linearly independent and thus may serve as a new basis of

$$S_r = \{(p_1, \ldots, p_{2^k-1})' \mid \sum_{i=1}^{2^k-1} p_i \leq 1, \text{ and } p_i \geq 0, i = 1, \ldots, 2^k - 1\}. \tag{A.9}$$

Since $\log f^{(r)}(\mathbf{p}_r)$ is concave, $\mathbf{p}_r$ maximizes $f^{(r)}$ if and only if along each direction $\boldsymbol{\delta}_i^{(r)}$,

$$\left.\frac{\partial f^{(r)}(\mathbf{p}_r + u\boldsymbol{\delta}_i^{(r)})}{\partial u}\right|_{u=0} = 0 \text{ if } p_i > 0; \quad \leq 0 \text{ otherwise.}$$

That is, $f_i(z)$ attains its maximum at $z = p_i$, for each $i = 1, \ldots, 2^k - 1$ (and thus for $i = 2^k$). Based on Lemma 7.1 and Lemma 7.2, it implies one of the two cases:

(i) $p_i = 0$ and $f_i\left(\frac{1}{2}\right) \cdot 2^{d+1} - f(\mathbf{p}) \leq f(\mathbf{p})(d+1)$;

(ii) $p_i > 0$, $a > b(d+1)$, and $a - b(d+1) = p_i(a-b)(d+1)$, where $b = f_i(0)$, and $a = \frac{f(\mathbf{p}) - b(1-p_i)^{d+1}}{p_i(1-p_i)^d}$.

The conclusion needed can be obtained by simplifying those two cases above. □

Proof of Theorem 3.2: Let $\mathbf{p}_I$ be the saturated design satisfying $p_{i_1} = p_{i_2} = \cdots = p_{i_{d+1}} = \frac{1}{d+1}$. Note that if $|X[i_1, i_2, \ldots, i_{d+1}]| = 0$, $\mathbf{p}_I$ can not be D-optimal. Suppose $|X[i_1, i_2, \ldots, i_{d+1}]| \neq 0$, $\mathbf{p}_I$ is D-optimal if and only if $\mathbf{p}_I$ satisfies the conditions of Theorem 3.1. By Lemma 3.1, $f(\mathbf{p}_I) = (d+1)^{-(d+1)}|X[i_1, i_2, \ldots, i_{d+1}]|^2 w_{i_1} w_{i_2} \cdots w_{i_{d+1}}$. For $i \in \mathbf{I}$, $p_i = \frac{1}{d+1}$, $f_i(0) = 0$. By case (ii) of Theorem 3.1, $p_i = \frac{1}{d+1}$ maximizes $f_i(x)$. For $i \notin \mathbf{I}$, $p_i = 0$,

$$\begin{aligned} f_i\left(\frac{1}{2}\right) &= [2(d+1)]^{-(d+1)}|X[i_1, \ldots, i_{d+1}]|^2 w_{i_1} \cdots w_{i_{d+1}} \\ &+ 2^{-(d+1)}(d+1)^{-d} w_i \cdot w_{i_1} \cdots w_{i_{d+1}} \sum_{j \in \mathbf{I}} \frac{|X[\{i\} \cup \mathbf{I} \setminus \{j\}]|^2}{w_j}. \end{aligned}$$

Then $p_i = 0$ maximizes $f_i(x)$ if and only if $f_i\left(\frac{1}{2}\right) \leq f(\mathbf{p})\frac{d+2}{2^{d+1}}$, which is equivalent to

$$\sum_{j \in \mathbf{I}} \frac{|X[\{i\} \cup \mathbf{I} \setminus \{j\}]|^2}{w_j} \leq \frac{|X[i_1, i_2, \ldots, i_{d+1}]|^2}{w_i}.$$

□

We need the lemma below for Theorem 5.2:

Lemma 7.3 *Suppose $k \geq 3$ and $d(d+1) \leq 2^{k+1} - 4$. For any index set $\mathbf{I} = \{i_1, \ldots, i_{d+1}\} \subset \{1, \ldots, 2^k\}$, there exists another index set $\mathbf{I}' = \{i'_1, \ldots, i'_{d+1}\}$ such that*

$$|X[i_1, \ldots, i_{d+1}]|^2 = |X[i'_1, \ldots, i'_{d+1}]|^2 \text{ and } \mathbf{I} \cap \mathbf{I}' = \emptyset. \tag{A.10}$$

Proof of Theorem 5.2: Fixing any row index set $I = \{i_1, \ldots, i_{d+1}\}$ of X such that $|X[i_1, i_2, \ldots, i_{d+1}]|^2 > 0$, among all the $(d+1)$-row fractional designs satisfying $p_i = 0, \forall i \notin \mathbf{I}$, $|X'WX|$ attains its maximum $\left(\frac{1}{d+1}\right)^{d+1} w_{i_1} \cdots w_{i_{d+1}} \cdot |X[i_1, i_2, \ldots, i_{d+1}]|^2$

at $\mathbf{p}_I$ satisfying $p_{i_1} = \cdots = p_{i_{d+1}} = \frac{1}{d+1}$. By Lemma 7.3, there exists an index set $\mathbf{I}' = \{i'_1, \ldots, i'_{d+1}\}$ such that $|X[i'_1, \ldots, i'_{d+1}]|^2 = |X[i_1, \ldots, i_{d+1}]|^2$ and $\mathbf{I} \cap \mathbf{I}' = \emptyset$. Let $\mathbf{w}_{I'} = (w_1, \ldots, w_{2^k})'$ satisfy $w_i = b, \forall i \in \mathbf{I}'$ and $w_i = a, \forall i \in \mathbf{I}$. Then the relative loss of efficiency of $\mathbf{p}_I$ with respect to $\mathbf{w}_{I'}$ is

$$R_{I'}(I) = 1 - \left(\frac{\psi(\mathbf{p}_I, \mathbf{w}_{I'})}{\psi(\mathbf{p}_{I'}, \mathbf{w}_{I'})} \right)^{\frac{1}{d+1}} = 1 - \frac{a}{b}.$$

Thus the maximum relative loss of efficiency of p_I is at least $1 - \frac{a}{b}$. If $\mathbf{I}$ maximizes $|X[i_1, \ldots, i_{d+1}]|^2$, then $\mathbf{p}_I$'s maximum relative loss of efficiency attains the minimum $1 - \frac{a}{b}$. □

We need two lemmas for the exchange algorithm for integer-valued allocations.

Lemma 7.4 *Let $g(z) = Az(m-z) + Bz + C(m-z) + D$ for real numbers $A > 0, B \geq 0, C \geq 0, D \geq 0$, and integers $m > 0, 0 \leq z \leq m$. Let Δ be the integer closest to $\frac{mA+B-C}{2A}$.*

(i) If $0 \leq \Delta \leq m$, then $\max_{0 \leq z \leq m} g(z) = mC + D + (mA + B - C)\Delta - A\Delta^2$ at $z = \Delta$.

(ii) If $\Delta < 0$, then $\max_{0 \leq z \leq m} = mC + D$ at $z = 0$.

(iii) If $\Delta > m$, then $\max_{0 \leq z \leq m} = mB + D$ at $z = m$.

Lemma 7.5 *Let $\mathbf{n} = (n_1, \ldots, n_{2^k})'$, $W_n = \mathrm{diag}\{n_1 w_1, \ldots, n_{2^k} w_{2^k}\}$, $f(\mathbf{n}) = |X' W_n X|$. Fixing $1 \leq i < j \leq 2^k$, let*

$$\begin{aligned} f_{ij}(z) &= f\left(n_1, \ldots, n_{i-1}, z, n_{i+1}, \ldots, n_{j-1}, m - z, n_{j+1}, \ldots, n_{2^k}\right) \\ &\stackrel{\triangle}{=} Az(m-z) + Bz + C(m-z) + D, \end{aligned} \tag{A.11}$$

where $m = n_i + n_j$. Then (i) $D > 0 \Longrightarrow B > 0$ and $C > 0$; (ii) $B > 0$ or $C > 0 \Longrightarrow A > 0$; (iii) $f(\mathbf{n}) > 0 \Longrightarrow A > 0$; (iv) $D = f(n_1, \ldots, n_{i-1}, 0, n_{i+1}, \ldots, n_{j-1}, 0, n_{j+1}, \ldots, n_{2^k})$. (v) Suppose $m > 0$, then $A = \frac{2}{m^2}\left(2f_{ij}\left(\frac{m}{2}\right) - f_{ij}(0) - f_{ij}(m)\right)$, $B = \frac{1}{m}\left(f_{ij}(m) - D\right)$, $C = \frac{1}{m}\left(f_{ij}(0) - D\right)$.

References

Atkinson, A. C., Donev, A. N. and Tobias, R. D. (2007). *Optimum Experimental Designs, with SAS*, Oxford University Press.

Chernoff, H. (1953). "Locally optimal designs for estimating parameters", *Annals of Mathematical Statistics* **24**, 586−602.

Cox, D. R. (1988). "A Note on Design when Response has an Exponential Family Distribution", *Biometrika*, **75**, 161−164.

Dorta-Guerra, R., González-Dávila, E. and Ginebra, J. (2008). "Two-level experiments for binary response data", *Computational Statistics and Data Analysis*, **53**, 196−208.

Fedorov, V.V. (1972). *Theory of Optimal Experiments*, Academic Press, New York.

González-Dávila, E., Dorta-Guerra, R. and Ginebra, J. (2007), "On the information in two-level experiments", *Model Assisted Statistics and Applications*, **2**, 173−187.

Graßhoff, U. and Schwabe, R. (2008). "Optimal design for the Bradley-Terry paired comparison model", *Statistical Methods and Applications*, **17**, 275−289.

Grimshaw, S. D., Collings, B. J., Larsen, W. A. and Hurt C. R. (2001). "Eliciting Factor Importance in a Designed Experiment", *Technometrics*, **43**, 133−146.

Hamada, M. and Nelder, J. A. (1997). "Generalized linear models for quality-improvement experiments", *Journal of Quality Technology*, **29**, 292−304.

Khuri, A. I., Mukherjee, B., Sinha, B. K. and Ghosh, M. (2006). "Design issues for generalized linear models: A Review", *Statistical Science*, **21**, 376−399.

Martin, B., Parker, D. and Zenick, L. (1987). "Minimize slugging by optimizing controllable factors on topaz windshield molding", In: *Fifth Symposium on Taguchi Methods, American Supplier Institute, Inc., Dearborn, MI*, 519−526.

McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models*, Second Edition. Chapman and Hall/CRC, Boca Raton.

Miller, A. and Sitter, R. R. (2001). "Using the Folded-Over 12-Run Plackett-Burman Design to Consider Interactions", *Technometrics*, **43**, 44−55.

Muller, W. G. (2007). Collecting Spatial Data: Optimum Design of Experiments for Random Fields, 3rd Edition, Spinger.

Nair, V., Strecher, V., Fagerlin, A., Ubel, P., Resnicow, K., Murphy, S. A., Little, R., Chakraborty, B. and Zhang, A.J. (2008). "Screening experiments and the use of fractional factorial designs in behavioral intervention research", *American Journal of Public Health*, **98**, 1354−1359.

Nocedal, J. and Wright, S. J. (1999). *Numerical Optimization*. Springer, New York.

Rao, C. R. (1973). Linear Statistical Inference and Its Applications, John Wiley & Sons, New York.

Seo, D. M., Goldschmidt-Clermont, P. J. and West, M. (2007). "Of Mice And Men: Sparse Statistical Modeling in Cardiovascular Genomics", *Annals of Applied Statistics*, **1**, 152−178.

Severin, V. (2000), "Comparing statistical efficiency and respondent efficiency in choice experiments", *Ph.D. Thesis, University of Sydney*.

Street, D. J. and Burgess, L. (2007). "The Construction of Optimal Stated Choice Experiments", *Wiley Series in Probability and Statistics*.

Stufken, J. and Yang, M. (2012). "Optimal Designs for Generalized Linear Models", In: *Design and Analysis of Experiments*, Volume 3: Special Designs and Applications, K. Hinkelmann (ed.), Wiley, New York.

Vasandani, V. and Govindaraj, T. (1995). "Knowledge Organization in Intelligent Tutoring Systems for Diagnostic Problem Solving in Complex Dynamic Domains", *IEEE Transactions on Systems, Man and Cybernetics*, 1076−1096 .

Wynn, H.P. (1970). "The sequential generation of D-optimum experimental designs", *Annals of Mathematical Statistics*, **41**, 1655−1664.

Yang, J., Mandal, A. and Majumdar, D. (2011). "Optimal designs for two-level factorial experiments with binary response", *Statistica Sinica*, to appear.

Zangwill, W. (1969), "Nonlinear Programming: A Unified Approach", Prentice-Hall, New Jersey.

Zayats, N. and Steinberg, D. M. (2010). "Optimal Design of Experiments When Factors Affect Detection Capability", *Pakistan Journal of Statistics*, **26**, 15−37.

Supplementary Materials

Additional Results for Example 4.1: Consider a 2^3 main-effects model with logit link. Suppose $\beta_1 = 0$. As a corollary of Theorem 4.1, the regular fractions $\{1, 4, 6, 7\}$, $\{2, 3, 5, 8\}$ are D-optimal half-fractions if and only

$$4\ \nu\left(|\beta_0| + |\beta_2| + |\beta_3|\right) \geq \nu\left(|\beta_0| + |\beta_2| + |\beta_3| - 2\max_{0\leq i\leq 3}|\beta_i|\right).$$

Note that $\nu(\eta) = \frac{1}{2+e^{\eta}+e^{-\eta}}$ for logit link, which is symmetric about 0. To simplify the notations, let $\beta_{2\vee 3} = \max\{|\beta_2|, |\beta_3|\}$ and $\beta_{2\wedge 3} = \min\{|\beta_2|, |\beta_3|\}$. The regular fractions $\{1, 4, 6, 7\}$, $\{2, 3, 5, 8\}$ are D-optimal half-fractions if and only if one of three

conditions below is satisfied:

$$\begin{aligned}
&\text{(i)} \quad |\beta_2| + |\beta_3| \le \log 2; \qquad\qquad (S.10)\\
&\text{(ii)} \quad |\beta_2| + |\beta_3| > \log 2,\ \beta_{2\vee 3} \le \log\left(1 + e^{-\beta_{2\wedge 3}} + \left[1 + e^{-\beta_{2\wedge 3}} + e^{-2\beta_{2\wedge 3}}\right]^{1/2}\right),\\
&\qquad \text{and } |\beta_0| \le \log\left(\frac{2\exp\{|\beta_2| + |\beta_3|\} - 1}{\exp\{|\beta_2| + |\beta_3|\} - 2}\right);\\
&\text{(iii)} \quad \beta_{2\vee 3} > \log\left(1 + e^{-\beta_{2\wedge 3}} + \left[1 + e^{-\beta_{2\wedge 3}} + e^{-2\beta_{2\wedge 3}}\right]^{1/2}\right),\ |\beta_{2\vee 3}| \le \log\left(\frac{2e^{|\beta_{2\wedge 3}|} - 1}{e^{|\beta_{2\wedge 3}|} - 2}\right),\\
&\qquad \text{and } |\beta_0| \le \log\left(\frac{2e^{\beta_{2\vee 3}} - 1}{e^{\beta_{2\vee 3}} - 2}\right) - \beta_{2\wedge 3}.
\end{aligned}$$

The above result is displayed in the right panel of Figure 2. In the x- and y-axis, we have plotted β_2 and β_3 respectively. The rhomboidal region at the center (marked as ∞) represents the region where the regular fractions will always be D-optimal, irrespective of the values of β_0. The contours outside this region are for the upper bound of $|\beta_0|$. Regular fractions will be D-optimal if the values of $|\beta_0|$ will be smaller than the upper bound with β_2 and β_3 falling inside the region outlined by the contour.

More proofs

Proof of Theorem 4.1: Given $\beta_1 = 0$, we have $w_1 = w_5 = \nu(\beta_0+\beta_2+\beta_3)$, $w_2 = w_6 = \nu(\beta_0+\beta_2-\beta_3)$, $w_3 = w_7 = \nu(\beta_0-\beta_2+\beta_3)$, $w_4 = w_8 = \nu(\beta_0-\beta_2-\beta_3)$. The goal is to find a half-fraction $\mathbf{I} = \{i_1, i_2, i_3, i_4\}$ which maximizes $s(\mathbf{I}) := |X[i_1, i_2, i_3, i_4]|^2 w_{i_1} w_{i_2} w_{i_3} w_{i_4}$. For regular half-fractions $\mathbf{I} = \{1, 4, 6, 7\}$ or $\{2, 3, 5, 8\}$, $s(\mathbf{I}) = 256 w_1 w_2 w_3 w_4$. Note that $|X[i_1, i_2, i_3, i_4]|^2 = 0$ for 12 half-fractions identified by $1 = \pm A$, $1 = \pm B$, $1 = \pm C$, $1 = \pm AB$, $1 = \pm AC$, or $1 = \pm BC$; and $|X[i_1, i_2, i_3, i_4]|^2 = 64$ for all other 56 cases.

Without any loss of generality, suppose $w_1 \ge w_2 \ge w_3 \ge w_4$. Note that the half-fraction $\{1, 5, 2, 6\}$ identified by $1 = B$ leads to $s(\mathbf{I}) = 0$. Then the competitive half-fractions consist of both 1 and 5, one element from the second block $\{2, 6\}$, and one element from the third block $\{3, 7\}$. The corresponding $s(\mathbf{I}) = 64 w_1^2 w_2 w_3$. In this case, the regular fractions are optimal ones if and only if $4w_4 \ge w_1$. □

Proof of Lemma 7.3: Note that $k \ge 3$ and $d(d+1) \le 2^{k+1} - 4$ imply $d + 1 \le 2^{k-1}$ and $\frac{d(d+1)}{2} < 2^k - 1$. Fix an arbitrary index set $\mathbf{I} = \{i_1, \ldots, i_{d+1}\} \subset \{1, \ldots, 2^k\}$. It can be verified that there exists a nonempty subset $\mathbf{J} \subset \{1, 2, \ldots, k\}$, for example $\mathbf{J} = \{1, 2\}$, such that *(i)* the rows $i_1, \ldots, i_{d+1}$ of $[-A_1, -A_2, A_3, \ldots, A_k]$ are the same as the rows $i'_1, \ldots, i'_{d+1}$ of $[A_1, A_2, A_3, \ldots, A_k]$, where $A_1, \ldots, A_k$ are columns of X corresponding to the main effects; *(ii)* $\mathbf{I}' = \{i'_1, \ldots, i'_{d+1}\}$ satisfies conditions (A.10). Actually, $|X[i_1, \ldots, i_{d+1}]|^2 = |X[i'_1, \ldots, i'_{d+1}]|^2$ is guaranteed by the construction of $\mathbf{I}'$. If $\mathbf{I} \cap \mathbf{I}' \ne \emptyset$, there exist two rows of $\mathbf{I}$, for example i_1, i_2, such that the two rows have same entries at $A_3, \ldots, A_k$ columns and different entries at A_1, A_2 columns. Note that

the row pair (i_1, i_2) corresponds to a unique subset $\mathbf{J}$. There are $2^k - 1$ possible $\mathbf{J}$ but only $\frac{d(d+1)}{2}$ possible pairs. Since $\frac{d(d+1)}{2} < 2^k - 1$, there is at least one $\mathbf{J}$ such that there is no pair of rows corresponding to it. That is, $\mathbf{I} \cap \mathbf{I}' = \emptyset$ for such a $\mathbf{J}$.

Exchange algorithm for real-valued allocations

Lemma 9.6 *Let $g(z) = Az(e-z) + Bz + C(e-z) + D$ for nonnegative constants A, B, C, D, e. Define $\Delta = \frac{eA+B-C}{2A}$.*

(i) If $0 \le \Delta \le e$, then $\max_{0 \le z \le e} g(z) = eC + D + \frac{(eA+B-C)^2}{4A}$ at $z = \Delta$.

(ii) If $\Delta < 0$, then $\max_{0 \le z \le e} = eC + D$ at $z = 0$.

(iii) If $\Delta > e$, then $\max_{0 \le z \le e} = eB + D$ at $z = e$.

Lemma 9.7 *Let $\mathbf{p} = (p_1, \ldots, p_{2^k})'$, $f(\mathbf{p}) = |X'WX|$, and*

$$\begin{aligned} f_{ij}(z) &:= f\left(p_1, \ldots, p_{i-1}, z, p_{i+1}, \ldots, p_{j-1}, e-z, p_{j+1}, \ldots, p_{2^k}\right) \\ &\stackrel{\triangle}{=} Az(e-z) + Bz + C(e-z) + D, \end{aligned}$$

where $1 \le i < j \le 2^k$ and $e = p_i + p_j$. Then (i) $D > 0 \Longrightarrow B > 0$ and $C > 0$; (ii) $B > 0$ or $C > 0 \Longrightarrow A > 0$; (iii) $f(\mathbf{p}) > 0 \Longrightarrow A > 0$; (iv) $D = f(p_1, \ldots, p_{i-1}, 0, p_{i+1}, \ldots, p_{j-1}, 0, p_{j+1}, \ldots, p_{2^k})$; (v) Suppose $e > 0$, then $A = \frac{2}{e^2}\left(2f_{ij}\left(\frac{e}{2}\right) - f_{ij}(0) - f_{ij}(e)\right)$, $B = \frac{1}{e}\left(f_{ij}(e) - D\right)$, $C = \frac{1}{e}\left(f_{ij}(0) - D\right)$.

Exchange algorithm for maximizing $f(\mathbf{p}) = f(p_1, \ldots, p_{2^k}) = |X'WX|$

1° Start with an arbitrary design $\mathbf{p}^{(0)} = (p_1^{(0)}, \ldots, p_{2^k}^{(0)})'$ such that $f(\mathbf{p}^{(0)}) > 0$.

2° Set up a random order of (i, j) going through all pairs

$$\{(1,2), (1,3), \ldots, (1, 2^k), (2,3), \ldots, (2^k - 1, 2^k)\}.$$

3° For each (i, j), if $e := p_i^{(0)} + p_j^{(0)} = 0$, let $\mathbf{p}^{(1)} = \mathbf{p}^{(0)}$ and jump to 5°. Otherwise, let

$$\begin{aligned} f_{ij}(z) &= f\left(p_1^{(0)}, \ldots, p_{i-1}^{(0)}, z, p_{i+1}^{(0)}, \ldots, p_{j-1}^{(0)}, e-z, p_{j+1}^{(0)}, \ldots, p_{2^k}^{(0)}\right) \\ &= Az(e-z) + Bz + C(e-z) + D \end{aligned}$$

with nonnegative constants A, B, C, D determined by Lemma 9.7.

4° Define $\mathbf{p}^{(1)} = \left(p_1^{(0)}, \ldots, p_{i-1}^{(0)}, z_*, p_{i+1}^{(0)}, \ldots, p_{j-1}^{(0)}, e - z_*, p_{j+1}^{(0)}, \ldots, p_{2^k}^{(0)}\right)'$ where z_* maximizes $f_{ij}(z)$ with $0 \le z \le e$ (see Lemma 9.6). Note that $f(\mathbf{p}^{(1)}) = f_{ij}(z_*) \ge f(\mathbf{p}^{(0)}) > 0$.

5° Repeat $2^\circ \sim 4^\circ$ until convergence (no more increase in terms of $f(\mathbf{p})$ by any pairwise adjustment).

Theorem 9.1 *If the exchange algorithm converges, the converged* $\mathbf{p}$ *maximizes* $|X'WX|$.

Proof of Theorem 9.1: Suppose the exchange algorithm converges at $\mathbf{p}^* = (p_1^*, \ldots, p_{2^k}^*)'$. According to the algorithm, $|X'WX| > 0$ at $\mathbf{p}^*$. Without any loss of generality, assume $p_{2^k}^* > 0$. Let $\mathbf{p}_r^* = (p_1^*, \ldots, p_{2^k-1}^*)$, $l_r(\mathbf{p}_r) = \log f_r(\mathbf{p}_r)$, and $f_r(\mathbf{p}_r) = f(p_1, \ldots, p_{2^k-1}, 1 - \sum_{i=1}^{2^k-1} p_i)$. Then for $i = 1, \ldots, 2^k - 1$, $\left.\frac{\partial l_r}{\partial p_i}\right|_{p_r^*} = \frac{1}{f(\mathbf{p}^*)} \cdot \left.\frac{\partial f_r}{\partial p_i}\right|_{p_r^*} = 0$, if $p_i^* > 0$; ≤ 0, otherwise. Thus $\mathbf{p}^*$ (or $\mathbf{p}_r^*$) locally maximizes $l(\mathbf{p})$ (or $l_r(\mathbf{p}_r)$), and $\mathbf{p}^*$ attains the global maximum of $f(\mathbf{p})$ on S. □

Similar to the lift-one algorithm, we may modify the exchange algorithm so that $\mathbf{p}^{(0)}$ won't be updated until all potential pairwise exchanges among p_i's have been checked. It can be verified that the modified exchange algorithm must converge.